ORAU

INDEPENDENT REVIEW OF THE DRAFT WORK PLAN–RADIOLOGICAL SURVEY AND SAMPLING, FEBRUARY 2018
FORMER HUNTERS POINT NAVAL SHIPYARD
SAN FRANCISCO, CALIFORNIA

General Comment: Overall, the plan provides adequate detail and includes necessary components of the further investigations planned at the site. However, there are technical issues that prevent the plan from forming a cogent approach between the planning and assessment phases of the radiological survey and site investigation process. These issues are manifested in the assumptions and methods used for determining the required number of samples, estimating scanning minimum detectable concentrations, and assessment of data for demonstrating compliance with the release criteria. The associated issues are designated as **Significant Comment** in the following section-specific comment matrix. Other issues have been identified as either a **Comment** for technical improvement or clarity, or as a **Minor Comment** when more editorial in nature.

Independent Review Comments				
Section	**Page**	**Paragraph**	Applicable Text	**Comment/Observation and Recommendation**
2	2-1	2nd	*In addition, workers in the onsite laboratory used a method to analyze radium (Ra)-226 that may have reported higher than actual radioactivity because of interference with naturally occurring uranium.* *TtEC presented CSMs in removal action completion reports that were based on potentially falsified data and screening results for Ra-226 reported by the onsite laboratory (results were often biased high).*	**Minor comment**: Suggest specifying that Ra-226 was quantified from the direct 186.2 keV photopeak, resulting in concentration overestimates due to interference of the U-235 185.7 photopeak in the spectrums region of interest. The recommended edit will provide additional technical clarifications and provide additional emphasis of the U-235 as a minor radionuclide of concern presented in Section 4 and for the naturally occurring radioactive material (NORM) evaluations discussed in Section 6.
2	2-5	Table 2-1	*Potential Source Areas/Release Areas:/Known Release Areas (from Page 6-38 of HRA); Potential Releases Identified after the HRA*	**Comment:** Although the information may be planned for inclusion in the task-specific plans (TSPs), this overarching work plan should include all germane information to its development and minimize information by referencing existing or future supplementary plans. Table 2-1 should include a brief discussion/description of the use of each building rather than referring to the Historical Radiological Assessment (HRA). Of particular note are the potential releases associated with Building 529. What was the source of the Cs-137 noted in the drain line sediments (~2,000 pCi/g)? Table 2.1 only refers to the removal of

Independent Review Comments				
Section	**Page**	**Paragraph**	Applicable Text	**Comment/Observation and Recommendation**
				contaminated sinks and drains. Which of the building uses could have impacted the drains, etc.? The HRA lists a potential point source buried behind it, but does not specify make-up of the point source (pg. 6-58 of HRA), as well as an isotope storage vault (Table 6-5A HRA), neutron generator, and low flux neutron lab and CW accelerator.
2	2-6	Table 2-1	*Potential Source Areas/Impacted Buildings*	**Comment**: As noted in the comment above, additional information regarding a summary of building use would be useful.
2	2-6	Table 2-1	*Radionuclides of Concern*	**Comment**: Related to use description comments, what is the justification for Th-232 only on building surfaces; U-235 only in Bldg. 365; Pu-239 for Bldg. 529 vault and drains, multiple building interiors; which buildings? The reader must assume that the information will be in a TSP; however, to the degree possible, the work plan should be a stand-alone document, particularly as it is the main driver behind what follows in TSP.
2	2-8	Table 2-1	Uncertainties *Non-impacted building had surveys performed at ends of pipes, and pipes were capped if no contamination was identified*	**Comment**: Because the text states "if no contamination was identified," an interpretation could be made that there were HRA-designated, non-impacted building pipes where contamination was identified, and, if so, the question must be asked for when the contamination was identified. Clarification of this uncertainty is necessary as the HRA is used to provide the basis behind the level of effort for the additional investigations.
4	4-11	2nd	*A site investigation will be conducted for the remaining radiologically impacted sites with lower contamination potential. The site investigation will entail a combination of soil sampling and judgmental scanning and static gamma measurements.*	**Minor Comment**: For consistency throughout the work plan, the text should specify that both random and judgmental soil sampling is planned.
4.1.1	4-12	Table 4.2	Tabulated *Project Release Criteria*	**Comment**: The table is not footnoted nor is any discussion provided in the plan that addresses whether the release criteria apply independently or if multiple radionuclides are present—criteria are collective.

Independent Review Comments				
Section	**Page**	**Paragraph**	Applicable Text	**Comment/Observation and Recommendation**
				The interpretation is that the criteria apply independently based on work plan content as no specific considerations are presented for multiple, comingled radionuclides. However, this may or may not be the correct interpretation. Footnotes within Table 1 of the 2006 Basewide Radiological Removal Action Memorandum identify limits for structures that are based on 25 mrem/year or Regulatory Guide 1.86, whichever is lower. Most of the respective annual doses provided in the Table are an order of magnitude less than the 25 mrem/year standard. The criterion for Th-232, indicated as a radionuclide of concern (ROC) on some buildings, equates to 25 mrem/year. As such, if other radionuclides are present with Th-232, the basic dose limit could conceivably be exceeded without the application of modified release criteria, such as application of the unity rule, modified criteria or gross activity criteria. This is not foreseen as issue for structures, as the work plan commits to applying the lowest of the criteria for data assessments, whereby all reported activity would be compared with the most conservative limit. The soil criteria show similar dose characteristics as described for structures; however, in the case of soil where radionuclide specific analyses are performed, the work plan should provide discussions or methods, particularly the unity, for planning and assessments. Otherwise, the document should clearly state that all criteria are applied independently.
4.1.2	4-12	1st	*Investigation levels are media-specific, radionuclide-specific concentrations or activity levels based on the release criteria that trigger the response.*	**Comment**: The noted statement may only be the case if the criteria are applied independently. There should be clear discussion as to how criteria are to be applied as discussed in the previous comment.

Independent Review Comments				
Section	**Page**	**Paragraph**	Applicable Text	**Comment/Observation and Recommendation**
4.1.2.1	4-12	1st	*The investigation level for gamma scan results will be established at three standard deviations above the mean for the gamma scan data set being evaluated.*	**Significant Comment**: The stated action level may not be appropriate unless it can be demonstrated that release criteria response would be greater than this value. Furthermore, this action level effectively only establishes a false positive limitation of 0.01 and increases the likelihood of false negative errors with the increased departure from the background distribution, and does not provide information on the desired true positive/false negative proportions. There are numerous other related comments to the proposed action level and establishment of the scanning minimum detectable concentrations (MDC_{SCAN}). A more defensible approach that provides greater control of decision errors is to establish a background threshold value based on other upper limit threshold value methods, such as 95^{th} percentiles, upper tolerance limits, etc. Additionally, with the planned five background areas, justifying the use of a specific background data set or pooled data sets may prove difficult.
4.1.2.1	4-12 and 4-13	Last/1st	*Gamma scan data will be evaluated using a posting plot. The posting plot will be used to identify areas of elevated activity. Areas of elevated activity will be further investigated by collecting biased samples.*	**Comment**: A technical basis is necessary to establish the methods for plotting and assessing geo-referenced data, including the background data compared. ORAU has presented multiple lessons learned on the inconsistent industry methods applied to presenting and interpreting geo-referenced scan data, as well as attempting to apply the NUREG-1507 MDC_{SCAN} calculation methods. Additional related comments are presented on the approach.
4.2	4-13	3rd	*Background values and standard deviations for soil and buildings will be determined to validate the number of samples required in each survey unit.*	**Comment**: It is unclear how using the background mean and standard deviations validates the number of samples required in each survey unit. If the Wilcoxon Rank Sum (WRS) test was planned, the higher of the values between the reference area and survey unit is used for calculating sample size. Is the intent of this statement to ensure that this practice is satisfied? Additionally, the methods used for determining sample size alluded to here and elsewhere in the plan (Section 4.3.3) are for estimating the mean

Independent Review Comments				
Section	**Page**	**Paragraph**	Applicable Text	**Comment/Observation and Recommendation**
				concentration in the survey unit, followed by a two-sample population test of the means/medians. The remainder of the plan primarily indicates that the background soil data will be used to NORM-related threshold values for comparison with individual survey unit results, without application of a hypothesis. As such, the plan lacks integration between the planning inputs for the number of samples—WRS test for comparing the survey unit average to the reference average plus a significant difference (release criteria)—and the intended data assessments, which are essentially threshold evaluations that are not connected.
4.2.1	4-16	1st	*The original samples will be split, so that one complete set will be available for analysis by another laboratory, if needed.*	**Comment**: Historically, ORAU's independent verification and analytical comparison shows that split samples many times introduce the question of when results do not agree. This then must be resolved by trading the samples between laboratories and comparing the results for both samples. Experience suggests it is best if the second lab analyzes the same sample when possible. The document should also provide information as how split sample data will be compared, such as via a duplicate error ratio (DER). Alternatively, if the information is presented in the Sampling and Analysis Plan/Quality Assurance Project Plan (SAP/QAPP), a reference should be provided.
4.3.1	4-17	All	*Soil Area Groups*	**Comment**: Concur with the plan to keep distinct soil populations as separate decision units (backfill vs. sidewalls/bottom). Based on this, the recommendation that the document would benefit if earlier in Section 4, Survey Design, perhaps a brief discussion or table were provided that listed primary decision units (PDUs), that is trench bottom and side wall undisturbed soil, various types of backfill soil, and further broken down by Group 1, 2a, and 2b. Figure 4.2 on pg. 4-18 somewhat accomplishes this, together with Sect. 4.3.2 on pg. 4-19 that addresses the concepts of PDUs, although this could be presented more concisely.

Independent Review Comments				
Section	**Page**	**Paragraph**	Applicable Text	**Comment/Observation and Recommendation**
4.3.3.	4-19	Table 4-3	*Table 4-3 Number of Samples in a Survey Unit*	**Observation**: In relation to the Section 4.2 comment on sample size, the calculation for delta and, ultimately, sample size is somewhat non-standard. Rather than subtracting the estimated mean concentration (LBGR) from the derived concentration guideline level (DCGL), background is shown as being subtracted, with the result showing that value input was 0, rather than a background concentration as indicated, or simply DCGL/σ. General guidance in the relative shift equation is that the LBGR be set to the estimated mean of the net concentration above background, then adjusting the LBGR up or down to achieve a reasonable size. If a hypothesis test were to be applied, there is a probability defined by the Type II error whereby a survey unit is not released that contains no added contamination. A prospective power curve was created for the inputs shown. The probability of a Type II error occurring is small even residual concentrations are at 75% of the DCGL, provided σ has not been underestimated. Ultimately, this shows a survey design with adequate power. Additional evaluations were performed using site data to assess the adequacy of the relative shift, primarily for the observed mean and σ for Cs-137. The relative of 1.8 was calculating using a mean = 0.02 and σ = 0.05, resulting in 28 samples. **Comment**: The actual comment for this section is not necessarily the inputs used to determine the required number of samples. Rather the issue is that the plan does not articulate how the required number of samples addresses the decision rules. The planning inputs are to determine the number of sample for estimating the mean concentration at a given confidence level. However, data assessments provided in Section 6 are a comparison of individual sample results with threshold values, based on the 95% upper confidence level of background. Planning inputs to investigate threshold value exceedances are very different than estimating mean concentrations in an area. Furthermore, comparing to a 95% upper confidence level (UCL) of background as threshold value is not the same as planning for a two-sample

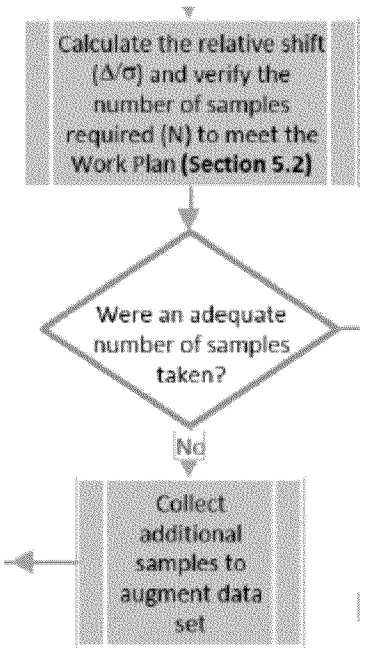
Independent Review Comments				
Section	**Page**	**Paragraph**	Applicable Text	**Comment/Observation and Recommendation**
				population test of the means for the survey unit and background plus a substantial difference. Generally, sampling plans that involve not-to-exceed threshold decisions are a probabilistic design whereby a random sample is collected from the total study area population units. For example, a 1,000 m^2 survey unit may be divided into 1 m^2 grids, or size of concern. Then the number of samples required is based on the desired confidence for detecting unacceptable units. Other plans are also available, based on both parametric and non-parametric upper tolerance levels. In all cases, the required number of samples is expected to exceed the 18 planned. Alternatively, the results for each survey unit, or collectively for a parcel, can be evaluated and an estimate of the percentage of the population likely to be acceptable at a stated confidence level.
4.4.1	4-20	1st	*Systematic and biased static measurements will be used to validate the survey unit classification before conducting alpha-beta scans to ensure that adequate scan coverage is achieved.*	**Comment**: Extensive radiological survey experience has shown that the probability is very low using random and even judgmental static measurements to identify potential contamination indicating survey unit misclassification. Qualitative surface scans combined with judgmental measurements typically are more efficient. These initial scans, if conducted to satisfy final status survey (FSS) data quality objectives, may either satisfy scan coverage requirements or provide the basis for increasing coverage. The recommendation is that this statement be replaced with some nominal scan coverage that is conducted first, perhaps 5 to 10% coverage representing both judgmental and systematic paths.
4.5	4.21		*The data quality objectives (DQOs) for the project are provided in Worksheet #11 of the SAP DQOs really belong in the work plan vs. the SAP.*	**Comment**: DQOs are a fundamental component of a work plan. The DQOs should be included in the work plan and precede most of Section 4 discussions.

Independent Review Comments				
Section	**Page**	**Paragraph**	Applicable Text	**Comment/Observation and Recommendation**
4.6.3	4-23	Table 4.5	[a] *Beta measurements are attributed to Sr-90, the most restrictive beta emitter (1,000 dpm/100 cm2 release criteria). For this reason, the instruments will be calibrated with 2π geometry for Sr-90 and yttrium-90.*	**Comment**: The Sr/Y-90 efficiency will be greater than if calibration were performed to a beta energy similar to Cs-137. Therefore, if the contamination is the result of a beta emitter, other than Sr/Y-90, surface activity will be underestimated when using Sr/Y-90 efficiency. As the Cs-137:Sr-90 release criteria is a factor of 5, the plan should provide information demonstrating that other release criteria would be also satisfied as the reduction in detection efficiency for other beta emitters of concern is much less than the factor of 5.
4.6.6.2	4-25	ALL	Calculation of Minimum Detectable Count Rates Calculation of Minimum Detectable Exposure Rate Calculation of MDCscan	**Significant Comment:** Derivation of the MDC_{SCAN} is not technically defensible due to multiple factors and is not integrated with specifications of investigation levels that have been stated to be based on count rates greater than the background mean+3σ. The following issues with the derivation were identified: 1. The plan specifies that the MDC_{SCAN} determinations were based on MARSSIM and NUREG-1507. However, the construct for deriving scan sensitivities in these references are based on human factors and decision making by listening to the detector audio output. The work plan does not indicate that surveyors will be trained to and required to listen to the audible response. Rather, all discussion for anomaly investigations appears to be specific to the assessment of post-processed data plots. These two constructs are very different, and the calculations in NUREG-1507 do not apply for post-processed data assessments. 2. The 6-second observation interval used in the b_i equation is not realistic nor is the 900 m^2 contaminated soil area modeled to determine the detector response. This size area corresponds to the area over which samples will be placed and the modeled method essentially equates to a static measurement or second stage scanning when pausing over an area. First stage scanning is intended to identify higher

Independent Review Comments				
Section	**Page**	**Paragraph**	Applicable Text	**Comment/Observation and Recommendation**
				activity anomalies that would not be identified by sampling and for which there is a brief observation interval. It is not appropriate to continue increasing the observation interval, akin to increasing the size of the *a priori* hot spot of concern, to lower the detection sensitivity. Furthermore the 0.4 pCi/g MDC_{SCAN} result is not realistic. Recently completed modeling determined a very conservative MDC_{SCAN} using a d' value of 2.32 for a 3-in × 3-in NaI detector of 1.8 pCi/g. 3. Cs-137 detection capability was not provided. 4. Related to Issue 1 above, the MDC_{SCAN} equation assumes a surveyor efficiency of 1. This is not an appropriate value to use simply because data are electronically captured. Human factors remain, unless mechanical methods are used, that will impact scan speed, surface-to-detector distance, pausing to allow the detector response to reach full scale, and other factors. As such, prior recommendations for similar cases regarding this issue were to set the surveyor efficiency to no greater than 0.75. Research continues in this area regarding surveyor efficiency vs. post-processed data evaluation efficiency for identifying anomalies and a desired true positive/false positive performance level.
5.2.1.1	5-2	2nd	*A minimum of 18 surface sample locations will be identified on a random-start systematic grid.*	**Minor Comment**: The December 2017 preliminary work plan file figures that were provided as examples showed rectangular grid, recommend triangular if the units are sized appropriately. Triangular spacing provides an increased probability of locating anomalies that may be present, in particular for subsurface areas or where the durable cover will not be removed and scan sensitivity is limited, at best.

Independent Review Comments				
Section	**Page**	**Paragraph**	Applicable Text	**Comment/Observation and Recommendation**
5.2.1.1	5-3	1st	*Gamma scan results exceed the investigation level and a potential for small areas of elevated activity have been identified will be investigated by collecting biased samples.*	**Comment**: Related to the >mean+3σ and MDC_{SCAN} determination comments, the plan should clearly state the intent or tie-together these different values into an integrated approach. Which investigation level is applicable?
5.2.2.2	5-5	1st	*Surface soil at former building sites and in crawlspaces underlying existing buildings in Group 2b areas will be surveyed as Class 3 survey units. Because these areas have a lower potential for radiological contamination, durable cover and pavement will not be removed from these areas. Because the durable cover or pavement effectively prevents most gamma radiation from escaping, gamma scan surveys of surfaces will not be performed*	**Comment**: Although detection efficiency will be extremely reduced, it is recommended that due diligence be implemented and to scan over the durable cover in light of the public concern with small commodity items. Experience has demonstrated at multiple sites that gamma-emitting commodities are many times detectable beneath these types of covers.
5.4.1	5-6	1st	*The number of data points, N, will be split between the survey unit and background reference area.*	**Comment**: Is there a specific reason for a structural background reference area and why would N be split between the two? Are the background reference areas for comparison in a two population hypothesis test, in which case generally an equal number of measurements are made in both populations, or is the intent to obtain construction material-specific background data to use in correcting survey unit gross counts to units of surface activity? Please clarify the intent.
6	6-1 and 6-3	5th bullet and Figure 6-2, 6-3	*The $DCGL_W$ test will be performed on the sample results.*	**Comment:** What is a $DCGL_W$ test? Is this simply a sample-by-sample comparison to the $DCGL_W$ or is a hypothesis test of the mean intended? Please clarify.

Independent Review Comments				
Section	**Page**	**Paragraph**	Applicable Text	**Comment/Observation and Recommendation**
6	6-1	3rd	*Once the laboratory data are available, verified, validated, and reviewed by a health physicist, comparisons to release criteria will be conducted. Survey units where all results exceed the 95 percent upper confidence level for the background reference area data set by a concentration less than the release criteria listed in Table 4-2 achieve the project objectives.*	**Significant Comment**: The statement, as written, is somewhat unclear. Is the following a correct interpretation of the second sentence? *Survey units* achieve the project objectives when *all results* are less than *the 95 percent upper confidence level for the background reference area data set* plus a significant difference equal to *the release criteria listed in Table 4-2.* Also the data quality assessment confidence statement is not consistent with prior confidence objectives discussed in the plan which is an example of another type of overall plan integration comment. Two confidence statements have been used thus far: the 95% UCL threshold and, in Section 4, sample planning was based on the 99% confidence level (Type I error of 0.01 in Section 4.3.3, Table 4.3)—not to be mistaken with the 99% upper confidence interval of the mean, which would result in a less conservative threshold value—that the mean of the survey unit was less than the $DCGL_W$. Additionally, information is lacking to confirm that there is no negative impact to the confidence and associated errors between designing the survey for 99% confidence that the survey unit mean is less than the DCGL via a WRS test with a background reference area sample population and instead, using the calculated 95% upper confidence level of the background reference area results to which the DCGL is added as the threshold. Please confirm the statistical relationships are the same.

Independent Review Comments				
Section	**Page**	**Paragraph**	Applicable Text	**Comment/Observation and Recommendation**
6.2	6-2	Figure 6-1	The right side of the flow chart, shown below, *NO* process flow from *Were an adequate number of samples taken?* decision box states "*Collect additional sample to augment data set*" 	**Significant Comment**: The intent of this step is unclear based on the threshold value decision basis. If parameter of demonstrating compliance were an estimate of the mean, with or without a hypothesis test, then additional samples are useful for reducing the uncertainty of the mean estimate. However, the understanding from the work plan review is that the decision is a binomial as to whether an individual sample does/does not exceed the background 95% UCL + significant difference. In addition, if the intent of the assessment is a hypothesis test, the flow chart suggests double sampling could be performed. The net result of double sampling is an increase in the Type I error of up to 2 times, which has not been addressed in the plan.
6.3	6-5	2nd	*The investigation level for gamma scan results will be established at three standard deviations above the mean for the gamma scan data set being evaluated.*	**Comment**: As discussed in prior comments, the planned investigation level and the MDC_{SCAN} derivation presented in Section 4.6.6.2 have not been integrated.
6.4.1	6-6	1st	*A background reference area data set will be developed and selected as representative of background for samples collected in a survey unit.*	**Comment**: Further details should be provided as to the selection of an appropriate background reference area data set for comparison with the survey unit.

Independent Review Comments				
Section	**Page**	**Paragraph**	Applicable Text	**Comment/Observation and Recommendation**
6.4.1	6-6	2nd	*Each sample result will be compared with the corresponding investigation level.*	**Comment**: Please clarify throughout the work plan the various decision methods described. Is there a relationship between the $DCGL_W$ test that is being referred to in the decision flow charts and Section 6 and the corresponding investigation level?
6.4.2	6-6	2nd	*The appropriate release criterion will be added to the mean for the background reference area data set selected to represent that survey unit. If all alpha static results are less than the alpha investigation level, and all beta static results are less than the beta investigation level, the survey unit meets the survey objectives.*	**Comment**: The discussion is somewhat unclear as to the intent of the assessment. Is this a discussion of a paired measurement hypothesis test or are the background reference data intended as construction material-specific backgrounds to be used in surface activity data conversions, which is a much simpler process to implement?
6.6.1 and 6.6.1.1	6-8	Figure 6-4	Calculate the NORM equation $e_n = s_r - (s_U * r)$ Where: e_n = NORM Evaluation Sr = sample nuclide activity (Ra-226 by emanation, U-235 & Th-232 by alpha Spec) s_u = sample U-238 activity (by alpha Spec) r = ratio of nuclide to U-238 from reference area data set d = nuclide $DCGL_W$ (release criteria)	**Comment**: An example calculation would be useful to clarify the application of the method and would assist in clarifying the discussion of the method in Section 6.6.1.1 The following questions on the calculation were noted: • How is r calculated for each radionuclide of concern? • Is it a mean ratio, a UCL ratio, an interval; is it from pooled background reference area data? • What if ratios are different between the background reference areas? • Would it be simpler to have the ratio as the threshold, and then develop the residual for comparison to DCGL? Also, please clarify if Ra, U-235 and Th-232 are assessed independently and if either fails, the sample fails.